Comparison of Two Probabilistic Techniques for the Assessment of Economic Uncertainty*

Dr. Dimitri N. Mavris
Assistant Professor & Manager
Aerospace Systems Design Laboratory (ASDL)

Mr. Oliver Bandte
Graduate Research Assistant, ASDL

School of Aerospace Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0150

Abstract

Several approaches to probabilistic design have been proposed in the past. Only few acknowledged the paradigm shift from performance based design to design for cost. The incorporation of economics in the design process, however, makes a probabilistic approach to design necessary, due to the inherent uncertainty of assumptions and the circumstances of operating environments of the future aircraft. The approach previously proposed by the authors, linking Response Surface Methodology with Monte Carlo Simulations, has revealed itself to be inadequate for multi-constraint, multi-objective problems. In addition accuracy problems were observed that could not be resolved with the methodology. Hence, this paper proposes an alternate approach to probabilistic design, which is based on a Fast Probability Integration (FPI) technique. The paper critically reviews the combined Response Surface Equation/ Monte Carlo Simulation methodology and compares it against the Advanced Mean Value (AMV) method, one of several Fast Probability Integration techniques. The Advanced Mean Value method is a probability estimation method based on a Most Probable Point (MPP) analysis. The paper describes the method employed to identify the Most Probable Point and obtain a cumulative probability distribution. The resulting distribution function is compared to the one generated by the Response Surface Equation/Monte Carlo Simulation method. For this comparison a case study is formulated, employing a High Speed Civil Transport concept. Based on the outcome of this study an assessment and comparison of the analysis effort and time necessary for both methods is performed. If the Most Probable Point can be found efficiently, the Advanced Mean Value method shows significant time savings over the Response Surface Equation/Monte Carlo Simulation method, and generally yields more accurate CDF distributions.

Introduction

The idea of a probabilistic approach to design is not a new one. Reliability engineering has used a probabilistic approach to structural design for many years.[1, 14, 27] It is only recently that these probabilistic methods found their way into the design of aerospace systems.[5, 7, 8, 11, 13, 16, 20, 21, 22, 29] Their sparse implementation is due in part to the fact that, for most design engineers the words 'uncertainty' or 'possibility' carry negative connotations, as in "the design is uncertain or imprecise." In addition, all 'legacy' design tools are based on inherently deterministic models, which cannot account for random changes of input values.

In recent years, systems design, in particular for aerospace systems, has experienced a paradigm shift from maximizing performance to maximizing affordability.[9, 19] Design for affordability requires the addition of cost estimation as a 'new' discipline to systems design and opens up the traditional deterministic approach to a probabilistic one. Many cost estimation methodologies are probabilistic in nature, since the assumptions made for the analysis are often uncertain.[9, 10] These methodologies treat the cost parameters as random variables and model their variation with probability distributions. This approach is suitable as long as the statistics of the various input variables are known. If this is not the case and their shape function or even range is unknown, a stochastic approach based on Fuzzy Logic is more suitable.[2] Modeling this kind of

* Paper presented at the 19th annual conference of the International Society of Parametric Analysts in New Orleans, May 1997

ambiguity through Fuzzy Logic is the basis of a subjective probability approach to economics and design. Nevertheless, the approach taken in this paper models all economic parameters as random variables with known shape function and range as a first approximation, leaving room for further development of the aerospace systems design methodology proposed in this paper. This paper critically reviews two approaches to probabilistic aerospace systems design, that allow for random changes in the aircraft operating environment and assumptions made in the design process.

Probabilistic Design in Economic Uncertainty Assessment

One of the major obstacles in applying probabilistic design methodologies is accommodating the large variety of computer codes used in modern systems design. It is impractical for all of them to be modified to accomodate a probabilistic problem formulation. Hence, a more generic methodology is proposed, which calls on some kind of 'wrapper' that, when linked to the analyses codes, drives the program and yields the desired results. Based on this formulation, probability functions can be defined for those input variables which are considered to be uncertain and a cumulative probability distribution function (CDF) for each of the desired objectives is obtained. Most probabilistic analyses, e.g. Monte Carlo Simulation [17], estimate their probability distribution functions based on a large number of samples generated over the design space, defined by the random variable ranges. While the usage of computer models allows for an easy perturbation of input values, an increase in complexity of the modeled system increases the complexity of the code and hence the run-time of the computer. Fox lists three methods that incorporate such complex computer programs in a probabilistic systems design approach.[11] Method #1, displayed in Figure 1, directly links a time consuming, due to a large number of repetitions needed, thus inefficient probabilistic method, such as the Monte Carlo Simulation, to the traditional systems design codes used in deterministic design approaches. Although computer speed has significantly increased in recent years, the extreme complexity of some design codes yields computation times that may prohibit a large number of program evaluations within the allotted time frame for the design process. Thus, Method #1 may not be a feasible option for a probabilistic design procedure.

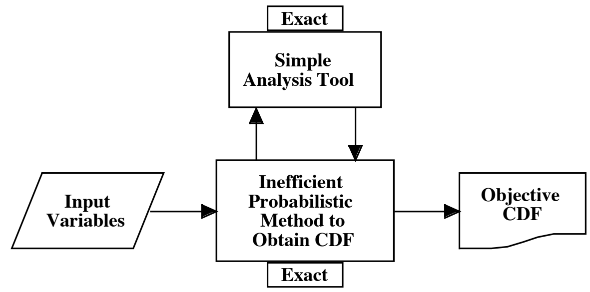

Figure 1: Probabilistic Design Method #1 [11]

Method #2, displayed in Figure 2, proposes the use of a metamodel which approximates exact design codes. The advantage of creating such a metamodel is a significantly reduced execution time, allowing a Monte Carlo Simulation to run on the metamodel rather than on the actual computer code. Several different metamodels have been proposed and applied. Some of the more common regression models are based on experimental designs [18], artificial neural networks [6], or Fuzzy Graph based metamodeling.[15]

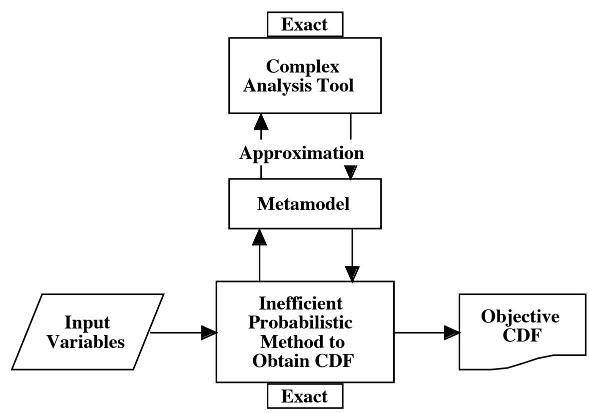

Figure 2: Probabilistic Design Method #2[11]

Method #3, displayed in Figure 3, takes a different approach, approximating the probability distribution function rather than the design code. This is based on the notion that in order to obtain the cumulative distribution function (CDF) not all probability levels need to be identified. The method selects several percentile levels and calculates the according objective value. Note that this calculation is based on the exact computer code, not on an approximating metamodel. These objective values and their probabilities can than be used to fit the typical S-shape of a CDF. The details of this method are described in [28] and in later sections of this paper.

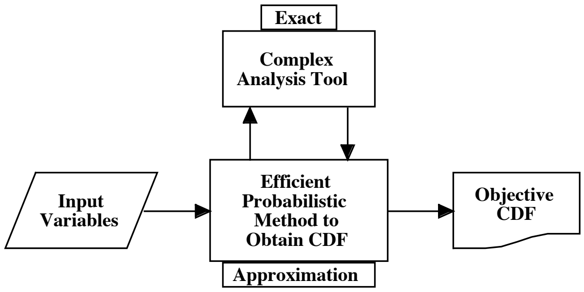

Figure 3: Probabilistic Design Method #3[11]

Method #2 has found the widest application and has also been used in the past by the authors.[20, 21, 22] In particular, the use of statistical regression models, based on Taylor series expansions, in combination with experimental designs is very popular.[5, 7, 8, 11, 13, 16, 29] The two main reasons for its popularity are its easy application to numerous computer simulation problems, e.g. aircraft synthesis, and the large number of statistical analysis tools commercially available, such as SAS(JMP), MINITAB, SPSS, etc. Nonetheless, there are two major problems in metamodeling of complex computer codes with a high number of inputs. First, the number of input variables handled by this approach is typically limited to eight or nine. This problem can often be solved through a screening process [3] that identifies the major contributors to variation in the model output. However, the metamodel created based on the screened parameters can never capture the variation of any of the other, ‘less important’ input parameters.

The second problem with Method #2 is in the mathematical background of such regression methods as Response Surface Methodology (RSM) and Design Of Experiments (DOE), which are based on random rather than deterministic variables (see [3], [4], [17], [18]). Therefore, the authors express caution in the straightforward application of these methods. Fundamental statistical knowledge is critical for obtaining reasonable approximations of the computer model. Many of the statistical analysis results the commercial packages offer are based on random error (ε) estimation and do not reflect the accuracy of the metamodel, since random error does not exist in deterministic computer simulation. A discussion on accuracy and behavior of statistical regression metamodels in computer simulations can be found in [18], [25], [26], and [30]. In general, the best validation of the accuracy of the metamodel is an extensive test at randomly distributed points over the design space to compare predicted values with the exact computer simulation values. Unfortunately, this test increases the computational effort put into the generation and validation of the metamodel. As shown in later sections of this paper, variation of only a subset of the variables in the metamodel can cause an additional prediction error not accounted for by testing the whole model.

Combined Response Surface/Monte Carlo Simulation Approach

Despite the aforementioned problems of the Response Surface Methodology (RSM), it can, if applied correctly, provide some valuable insight into the systems design code behavior. Hence, it has been used by the authors as a metamodel generator to facilitate probabilistic aerospace systems design methods.[21] In order to compare that approach (Method #2) with the one introduced in this paper (Method #3), a very brief overview of the combined Response Surface Equation/Monte Carlo Simulation (RSE/MCS) method is provided here. For more detailed information refer to References 3, 4, 20, and 21.

RSM is based on a statistical approach to build and rapidly assess empirical metamodels.[3, 4] By carefully designing and analyzing experiments or simulations, the methodology seeks to relate and identify the relative contributions of various input variables to the system response. However, modern aerospace systems are extremely complex, and most responses of interest are a function of many hundreds of design variables. The first step in constructing a Response Surface Equation (RSE) as a metamodel is to conduct a screening test to identify the variables which make the greatest contribution to the response of the system. The screening test is a two level fractional factorial Design of Experiments that accounts for main effects of variables only (i.e. no interactions).[4] It allows the rapid investigation of many variables to gain a first understanding of the problem.

After identifying the variables which will form the RSE, an experimental Design of Experiments has to be selected. For the purposes of this study, a face-centered central composite design was used as a scheme for the input variable levels to be tested. This experimental design is a three level composite design formed by combining a two-level full factorial with a star design.[3] Typically, a second order model in k-variables is assumed to exist. This second order polynomial for a response, R, can be written as:

$$R = b_o + \sum_{i=1}^{k} b_i x_i + \sum_{i=1}^{k} b_{ii} x_i^{\,2} + \sum \sum_{i<j}^{k} b_{ij} x_i x_j \quad (1)$$

where: b_i are regression coefficients for linear terms
b_{ii} are coefficients for pure quadratic terms
b_{ij} are coefficients for cross-product terms
x_i, x_j are the design variables of interest

Refer to References 3 and 4 for a detailed description of a response surface generation. After the RSE is developed, the effect of uncertain variables can be incorporated into a systems level design through the use of a Monte Carlo Simulation (MCS). A Monte Carlo Simulation is effectively a random number generator that selects values for each random variable with a frequency proportional to the shape of the corresponding probability distribution. Usually 5,000 to 10,000 trials are needed for a good representation of the response probability distribution. Without the aid of the RSE, this task would be computationally excessive and in many cases impractical considering that MCS would have to execute the design simulation code each time (Method #1, Figure 1).

Fast Probability Integration Approach

To avoid the often difficult generation of a metamodel (Method #2), the paper suggest the use of a fast probability integration technique as an approach to Method #3. This technique is reviewed here in greater detail. The Fast Probability Integration (FPI) computer program [28], developed by researchers at the Southwest Research Institute (SwRI) for the NASA Lewis Research Center, is a probability analysis code based on the Most Probable Point (MPP) analysis frequently used in structural reliability analysis. The MPP analysis utilizes a response function Z(X) that depends on several random variables X_i (see Figure 4 for a 2-D example). Each point in the design space spanned by the X's has a specific probability of occurrence according to their joint probability distribution function (see Figure 5). However, each point in the design space also corresponds to one specific response value Z(X). Hence, each response value has the same probability of occurrence as the corresponding point in the design space.

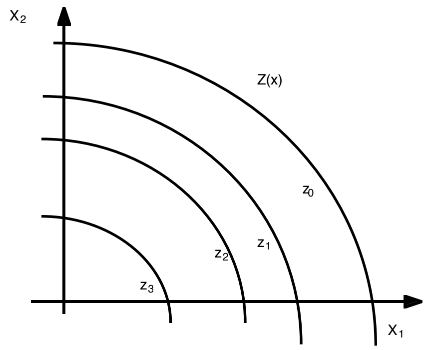

Figure 4 : Objective Function Contours

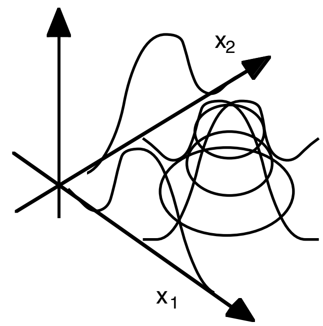

Figure 5 : Joint Probability Distribution

In cost analysis and other disciplines involving random variables, it is often desired to find the probability of achieving response values below a critical value of interest z_0. This critical value can be used to form a limit-state function (LSF):

$$g(X) = Z(X) - z_0 \quad (2)$$

where values of $g(X) \geq 0$ are undesirable. The MPP analysis calculates the cumulative probability of all points that yield $g(X) \leq 0$ for the given z_0 (see Figure 6). Since the LSF 'cuts off' a section of the joint probability distribution (see Figure 7) a point with maximal probability of occurrence can be identified on that LSF. This point is called the Most Probable Point. It is found most conveniently in a transformed space (see Figure 7), in which all random variables are normally distributed. Once the MPP and the cumulative probability are identified, the process can be repeated for several z_0 values mapping each probability over z_0. This cumulative probability distribution for Z(X) can than be differentiated to obtain the probability density function of the response.

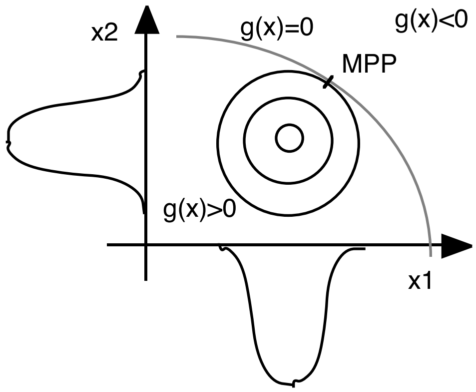

Figure 6: Most Probable Point (MPP) Location

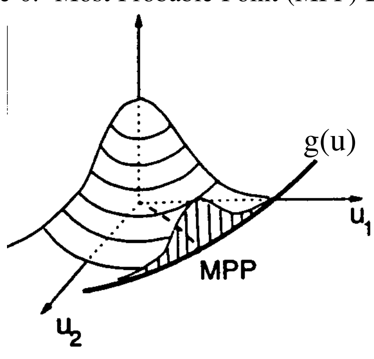

Figure 7: Visualization of MPP [28]

The FPI code offers several techniques to find the MPP and the probability of a given LSF value z_0 for the response function. Some of these techniques are very efficient and eliminate the need for an expensive Monte Carlo Simulation. An additional advantage of FPI is the fact that it is directly linked to the analysis code, eliminating the need for a metamodel and its limit in the number of variables. However, all Fast Probability Integration techniques approximate the LSF locally at the Most Probable Point.

Advanced Mean Value Method

The Advanced Mean Value (AMV) method is one of the twelve analysis methods in the FPI code. It combines a simple Mean Value method with the MPP analysis and determines the CDF for the response function Z(X). The Mean Value (MV) method is based on a simple Taylor series expansion of the response function Z(X) (Equation 3), assuming Z(X) to be smooth and the expansion to exist at the mean:

$$\begin{aligned} Z(X) &= Z(\mu) + \sum_{i=1}^{n} \left(\frac{\partial Z}{\partial X_i} \right) \cdot \left(X_i - \mu_i \right) + H(X) \\ &= a_o + \sum_{i=1}^{n} a_i X_i + H(X) \\ &= Z_{MV}(X) + H(X) \end{aligned} \tag{3}$$

The derivatives are evaluated at the mean values and $Z_{MV}(X)$ represents the sum of the first order terms and H(X) represent higher order terms. For n random variables, the a_i's can be estimated with n+1 function evaluations and a numerical differentiation method. Based on this linear approximation the CDF for $Z_{MV}(X)$ can be obtained directly, since the distributions for the random variables X_i are fully defined and $Z_{MV}(X)$ is explicit. For nonlinear Z-functions the MV solution for the CDF is not sufficiently accurate. One possibility for increasing accuracy is to increase the order of the Taylor series expansion, which becomes difficult and inefficient for implicit response functions and a large number of random variables (n).

A more efficient approach to increasing the accuracy is proposed by the AMV method:

$$Z_{AMV}=Z_{MV}+H(Z_{MV}) \tag{4}$$

$H(Z_{MV})$ is defined as the difference between Z and Z_{MV} at the Most Probable Point Locus (MPPL) of Z_{MV}, where the MPPL combines the MPP's for several values of z_0.[31] In other words, $H(Z_{MV})$ in Equation 4 approximates H(X) in Equation 3. Z_{AMV} would be exact if the MPPL was known and exact, i.e. $MPPL(Z_{MV}(X)) = MPPL(Z(X))$. Since the MPPL is not known, the AMV method approximates the locus based on Z_{MV}, which is for smooth response functions a good approximation.[28] Again, to avoid confusion with Method #2, the AMV method does not approximate the response function to obtain the CDF but rather the MPP (MV method). This approximation, however, is corrected by the move in the AMV method, as depicted in Figure 8. The steps for a CDF generation with the AMV method are also displayed in Figure 8.

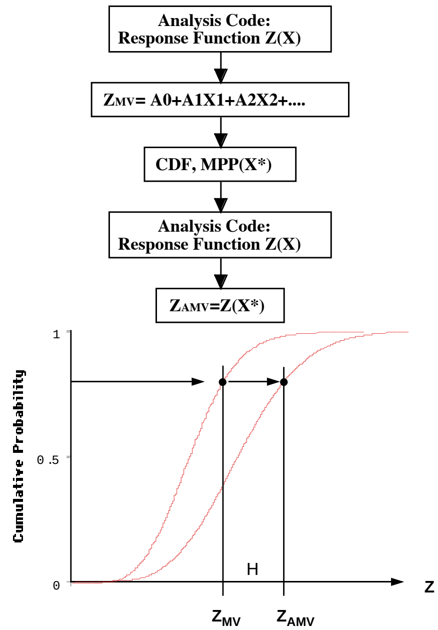

Figure 8: AMV Method [28]

One of the dominant advantages of the AMV method is the small number of function calls necessary. Where n+1 analysis code executions are sufficient for the linear approximation of the response function Z_{MV} and ten additional program evaluations are needed to obtain the updated Z_{AMV} for ten selected levels of z_0.[28] This translates into significant time savings over the RSE/MCS method which usually requires several hundred function evaluations for the generation of the RSE.[3] Additionally, the AMV is principally not limited to a small number of variables. The current limit within the FPI code of 100 variables is due to vector formatting and not the fast probability integration technique itself. Nonetheless, there is an additional gain associated with the extended effort in the RSE generation. It can serve as a valuable tool to gain understanding of the behavior of the underlying model. The AMV method, on the other hand, will only return a probability distribution without providing any further insight into the analysis code.

Comparison for an Example Equation

In order to get a first feel for the accuracy of the CDFs generated by both the RSE/MCS and AMV methods, Equation 5 is employed as an example simulation model that represents the complexity of the design tools used in aerospace systems design. For the RSE/MCS method, Equation 1 is fit to the data generated by a CCF design based on the ranges given in Table I. Equation 6 is the regressed quadratic equation, also displayed in form of prediction profiles in Figure 9. Unfortunately, the fit of the quadratic polynomial as a metamodel for the complex equation is only moderate, indicated by the whole model test in Figure 10 and the regression R^2 value of 0.985576. The residual plot in Figure 11 shows in particular the bad clustering of the data. The standard statistical analysis is insufficient for this type of data, since the underlying assumption of normally distributed prediction errors[4] is clearly rejected in Figure 11. However, in order to compare the two methods, the RSE will be used in the following Monte Carlo Simulation. The poor regression results just prove the point made earlier, that statistical regression metamodels should be used with caution.

$$Y = 5 + X_1 \cdot e^{X_2 + 3X_3} + 4X_4 \cdot e^{2X_5 - X_6} + X_7 + X_1 X_3 + X_2^2 \quad (5)$$

Table I: Ranges and Moments of Random Variables

Variable	Minimum	Maximum	Mean	Std.Deviation
X1	2	4	3	0.3
X2	1	3	2	0.3
X3	0.5	0.8	0.65	0.05
X4	3	6	4.5	0.3
X5	1	2	1.5	0.15
X6	1	4	2.5	0.4
X7	4	9	6.5	0.7

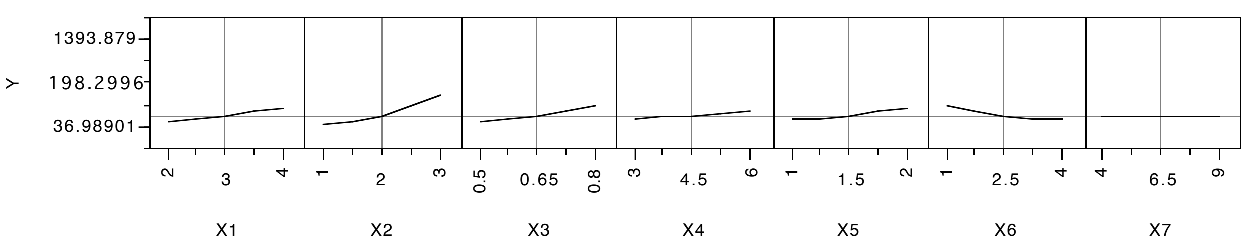

Figure 9: Prediction Profiles for Example Equation

$$\begin{aligned} Y = {} & 1381.4725 - 238.526X_1 - 728.0544X_2 - 2357.086X_3 - 14.99X_4 - 7.55X_5 + 48.53X_6 - 9.44X_7 \\ & + 5.02X_1X_1 + 67.32X_2X_1 + 90.635X_2X_2 + 249.62X_3X_1 + 568.04X_3X_2 + 936.13X_3X_3 + 2.23X_4X_4 \\ & + 36.464X_5X_4 + 84.546X_5X_5 - 14.44566X_6X_4 - 99.01554X_6X_5 + 20.07X_6X_6 + 0.803X_7X_7 \end{aligned} \quad (6)$$

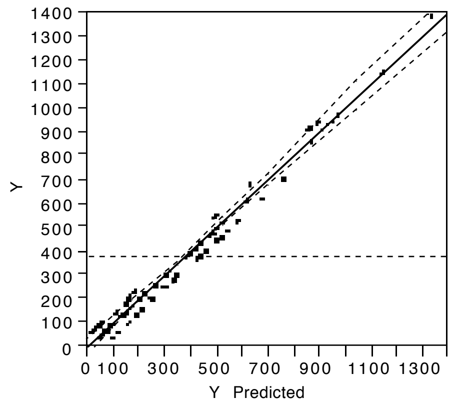

Figure 10: Whole Model Test

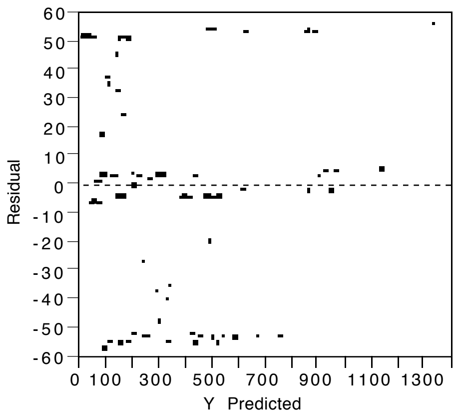

Figure 11: Residual Plot

The CDFs generated for the RSE/MCS and the AMV method are based on normal distributions for the seven variables with the statistical moments displayed in Table I. An overlay plot, Figure 12, depicts a clear discrepancy between the CDF based on the RSE/MCS method and the CDF from the AMV method. An additional CDF based on the actual equation using a Monte Carlo Simulation is displayed in Figure 12, to compare the two approximated with the 'true' CDF. This comparison is only feasible, because the model used for this study, Equation 5, is explicit. As the overlay plot depicts, the AMV method estimates the CDF much more accurately than the RSE/MCS method which can be contributed to the poor fit of the regression equation.

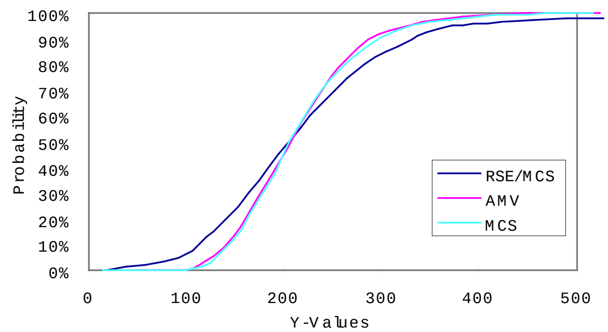

Figure 12: Cumulative Distribution Overlay Plot

Application Example: a High Speed Civil Transport

To further compare the two approaches to probabilistic design, an aerospace systems design example is examined in some detail here. The aircraft baseline used for this example is a High Speed Civil Transport (HSCT) depicted for review in Figure 13. The vehicle has an area-ruled fuselage (maximum diameter of 12 ft.), a double delta planform, and four nacelles below the wing housing mixed flow turbofan (MFTF) power plants. The values for some of the important design parameters are given in Table II. The mission profile for this aircraft is depicted in Figure 14, where the length of the subsonic cruise segment varies between 0 and 25 % of the design range. This split subsonic/supersonic mission is a result of the restriction for supersonic flight over land. In this study, the aerodynamics, structural arrangement, and engine cycle parameters are assumed to have all been optimized off-line and are held constant throughout this study. Thus, the aircraft in this case is only allowed to be scaled up or down to accommodate the assigned mission requirement variations.

Figure 13: HSCT Example

Table II: Description of the Baseline HSCT

Parameter	Baseline
Range	5000 nm
Payload	300 Passengers
Fuselage length	310 ft.
Span	77.5 ft.
Inboard Sweep	74 deg.
Outboard Sweep	45 deg.
Reference Area	9,000 ft^2
Mach Number	2.4
Cruise Altitude	~63,000 ft.
Sustained Load	2.5 g

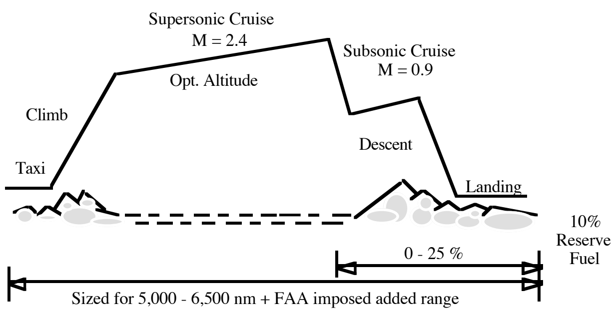

Figure 14: Baseline Mission Profile

For this study, the Flight Optimization System (FLOPS) [23] code was selected as the design simulation tool, while the Aircraft Life Cycle Cost Analysis (ALCCA) [12] program was selected as the economics model. Note, that the aircraft is sized for this demonstration of a probabilistic aerospace systems design without applying any external constraints, such as approach speed or landing field length. Based on previous screening tests performed by the authors [20, 21] an inclusive list of design and economic variables were identified and are listed in Table III as the main contributors to the response, which is the average required yield per revenue passenger mile ($/RPM).

Table III: Control and Noise Variable Descriptions and Ranges

		Variable	Range	Mean	Variance
Design Variable	Mission Characteristics	Design Range [nm]	5000....6500		
		% Subsonic Mission	0....25%		
	System Sizing	Wing Area [ft^2]	8500....9500		
		Thrust-to-Weight (T/W) Ratio	0.28....0.32		
		Number of Passenger (PAX)	220....320		
Uncertainty	Economics	Fuel Cost [$/gal]	0.55....1.10	0.75	0.07
		Load Factor	55....75%	0.65	0.04
		Utilization [hrs/year]	4500....5500	5000	200
		Economic Range [nm]	3000....5000	4000	350

A distinction is made between the Economic and the Design Range. Where the former represents the average distance a airplane will fly from one airport to another during its life, the latter depicts the maximum distance the aircraft is designed to fly. The nine parameters have been characterized as either design or uncertainty variables, where the uncertainty variables are associated with normal probability distributions. The design variables are not random but rather assumed to be under the control of the designer. Note that all randomness in this study is inherent in the economic uncertainty.

As part of the Response Surface Methodology, a face centered Central Composite Design [3] is identified for the nine variables in Table III. The typically assumed second order regression model, Equation 1, is used to estimate the relationship of design and economic variables with the response. Using the obtained RSEs, prediction profiles, depicted in Figure 16, can show the individual dependency or sensitivity of the response to the design and economic variables. All sensitivities are displayed for the baseline aircraft as the variable settings indicate. The random variables are set at their mean values. It should also be mentioned that throughout the study all actual values for $/RPM have been reduced by a constant to protect any sensitive data.

The R^2 value for the regression is 0.995161, indicating a successful fit of the data generated by the CCF design to the model in Equation 1. This value, however, does not reflect the prediction performance of the equation at 'off design' points. To verify the prediction accuracy of the RSE, 500 data points randomly distributed over the design space were generated, both for the actual design code and the RSE. The correlation plot for these 500 points, Figure 15, and a correlation R^2 value of 0.9958 indicate a very good prediction performance of the RSE. Indeed, the maximum prediction error found in the 500 sample points is only 4.22%.

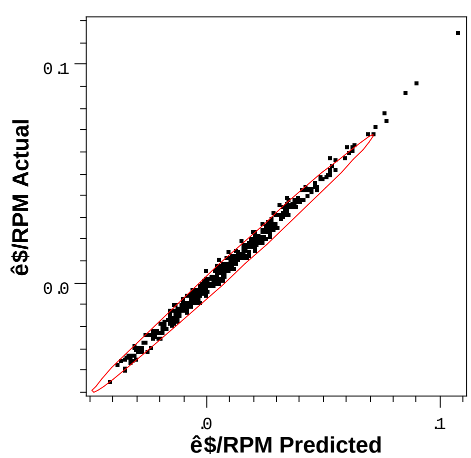

Figure 15: Correlation Plot

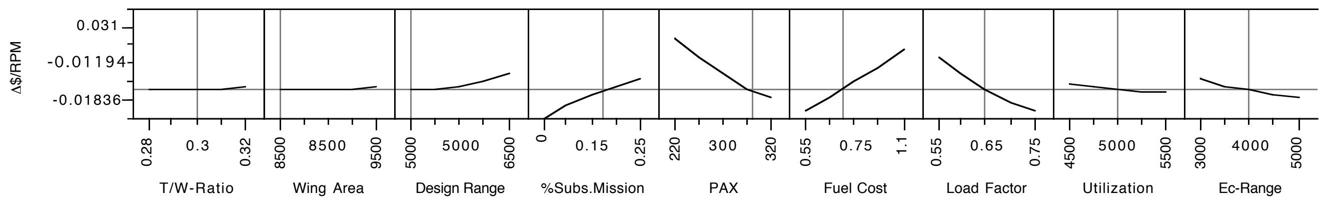

Figure 16: Response Surface Prediction Profiles

Comparison of CDFs for RSE/MCS and AMV Approach

Based on the previous example, Equation 5, the AMV method clearly yields a better estimate of the CDF, which can be contributed to the poor fit of the regressed metamodel. The objective becomes now the identification of any prediction differences between the AMV and the RSE/MCS method for the HSCT case study. Since the regression and validation of the quadratic metamodel was very successful, a good prediction can be expected from the RSE/MCS method. However, the RSE will be used in this study in a slightly different fashion. In the earlier example, all variables (X1 through X7) were used as random variables for the CDF generation, while for the HSCT case study only four of the nine variables are random. The RSE for the HSCT was originally generated for a robust optimization procedure [21], which identifies the design solution that minimizes the dependency of the response on the random variable variation. Thus, the RSE for the objective function, $/RPM, is a function of both random and design variables. This robust design optimization problem is not executed in this paper, although the herein proposed approach to probabilistic design can be easily applied to the Robust Design Simulation method proposed in [22].

Since the design variable settings are fixed for the CDF generation, all variation in the probability distribution are contributed to the variation of the uncertain economic variables. The design variables, however, affect the mean of the distribution significantly, exhibited for the different means of the distributions in Figure 17 and Figure 18. The CDFs shown in Figure 17 corresponds to the midpoint

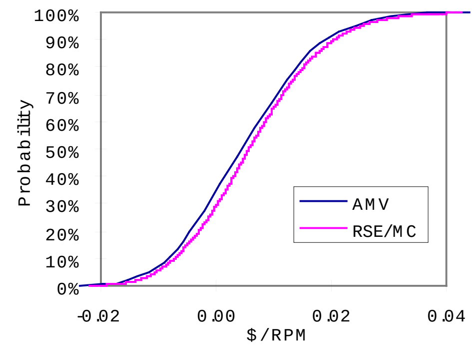

Figure 17: CDF Comparison for the Midpoint

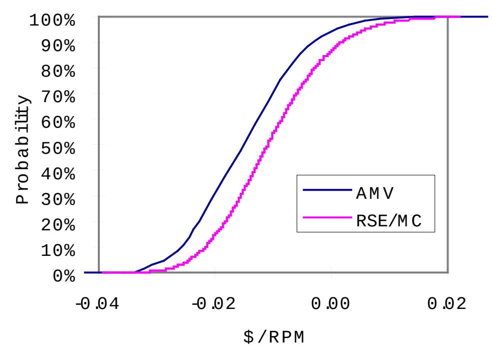

Figure 18: CDF Comparison for the Baseline Configuration

values of the design variables, while the CDFs in Figure 18 are based on the baseline configuration in Table II and Figure 15. Based on the study for the example equation, the CDF produced by the AMV method is assumed to be closest to the 'true' CDF, generated by a Monte Carlo Simulation.

It is not necessarily intuitive to contribute the noticeable difference in mean of the CDFs in Figures 17 and 18 to the prediction error of the RSE, which is rather small as demonstrated before. Nevertheless, the

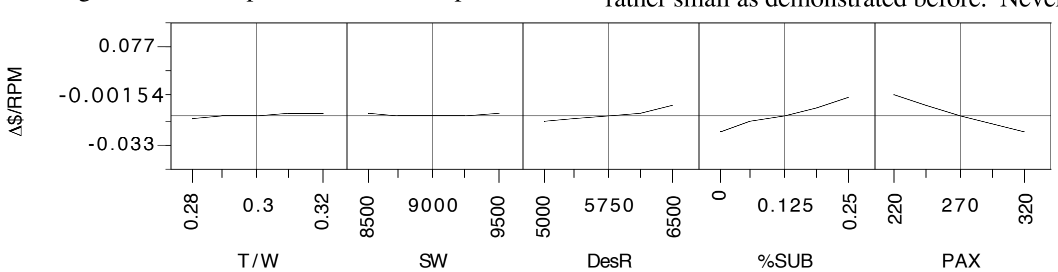

Figure 19: Prediction Profile for Design Variables RSE

design variables are deemed to be responsible for the prediction error, since for the same random variable distributions the CDF differs with the setting of the design variables. To investigate this phenomenon an additional RSE is created, this time for the five design variables only, keeping the economic variables constant at their mean values. This RSE is displayed in form of prediction profiles in Figure 19. The regression fit for this quadratic equation is surprisingly worse than the one for the nine variable RSE before, indicated by the lower R^2 value of 0.984062. This result in itself surprises, since in theory the regression fit gets worse with an increase in the number of variables due to the increase in complexity of the surface to be fitted.[24] An explanation for this regression behavior can be found in the nature of the analysis codes used for this study. While the economics package, ALCCA, is based on simple regression curves and smooth relationships between inputs and outputs, FLOPS, the synthesis and sizing tool used for this study, incorporates table lookup routines and internal iteration loops that create unsmooth surfaces. The nine variable RSE, however, is based on inputs to both programs, while the five variable RSE is based only on inputs to FLOPS. Apparently the four economic variables have the tendency to smooth out the complex surface created by the design variables. Unfortunately, a visualization of this effect is difficult to produce with the statistical analysis tools at hand.

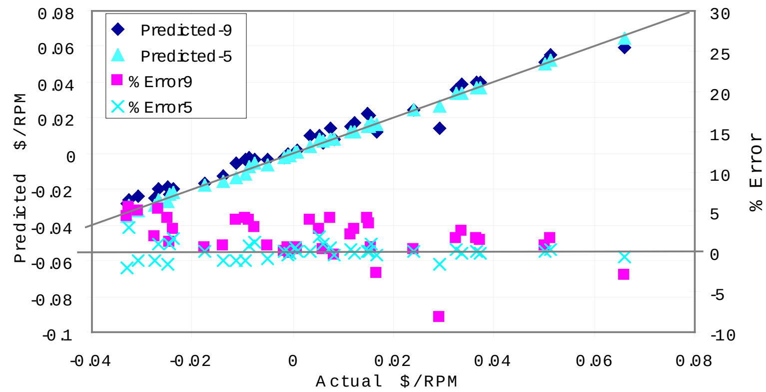

Figure 21: Prediction Error of the Nine Variables RSE vs. Five Variables RSE

Table IV: Prediction Error for Nine and Five Variable RSEs

	Δ$/RPM Error 9 Variable RSE	% Error 9 Variable RSE	Δ$/RPM Error 5 Variable RSE	% Error 5 Variable RSE
Baseline Config.	-0.0123	-1.4	-0.0163	1.353
Mid Point	0.00524	0.1	0.0056	-0.125
Minimum Error	-0.00189	-0.0013	-0.00185	-0.027
Maximum Error	0.014	8.1	0.186193	1.584

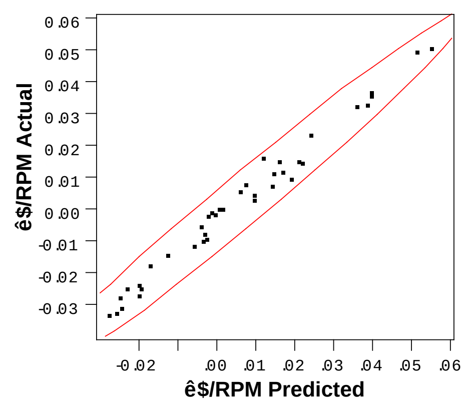

Figure 20: Correlation Plot for 42 Test Cases

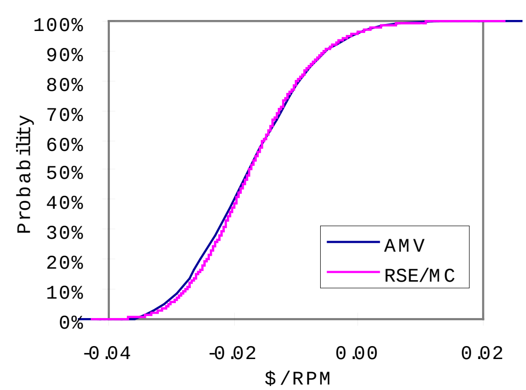

Figure 22: CDF Comparison for a Point With Minimal Error

Based on the previous argument, most of the prediction error of the nine variable RSE is compounded in the prediction of the mean of the CDF. To estimate this prediction error the nine variable RSE is tested at the 42 design points generated for the five variable CCF. A correlation graph of these 42 points for the actual and the predicted Δ$/RPM values is presented in Figure 20. It can be concluded from this graph and an R^2 value of 0.963 that the prediction performance of the nine variable RSE, with the design variables only, is rather poor, yielding a significantly different CDF in the probabilistic analysis. Table IV presents selected design points and their corresponding prediction errors for both RSEs. For additional illustration, the graph in Figure 21 maps the predicted values from both RSEs and their corresponding error over the actual values of Δ$/RPM. It can be seen from Figure 21 that the five variable RSE fits the data very well, which is expected, since the RSE was fit to this data set. The nine variable RSE, on the other hand, consistently overpredicts the $/RPM values, which explains the shift of the mean in Figures 17 and 18. Finally, for a minimal prediction error a CDF can be presented, Figure 22, that matches the CDF based on the AMV method almost exactly.

Conclusions

The present paper reviewed three different concepts to probabilistic design, which were executed and compared to each other. Method #1 applies a Monte Carlo Simulation directly to the analysis codes used in the design process. This approach yields an accurate estimate of the objective probability distribution, but requires a large number of code evaluations, which is usually too time consuming for most modern systems design tools. Method #2 requires the generation of an approximate metamodel to facilitate a Monte Carlo Simulation in a more time efficient manner. Method #3 approximates the probability distribution rather than the design simulation code, extracting the desired probabilistic information based on a significantly reduced number of function calls. Methods #2 and #3 were compared against each other on the basis of two example studies, one employing a complicated exponential equation, the other using a High Speed Civil Transport concept as an aerospace systems design example. An experimental design was used as an approach to Method #2, generating a Response Surface Equation which was employed in a Monte Carlo Simulation. Method #3 executed the Advanced Mean Value method, which is based on a fast probability integration technique. The generated cumulative distribution functions from both methods were compared against each other, determining that for both examples the Advanced Mean Value method yields the more accurate estimate of the probability distribution. Considering also the reduced number of function calls necessary for the analysis and the ability to accommodate more variables, the authors conclude the Advanced Mean Value method to be the more efficient and effective approach to probabilistic design.

References

[1] Ang, A. H.-S., Tang, W. H., Probability Concepts in Engineering Planning and Design, Volume I+II John Wiley & Sons, Inc., New York, NY, 1984.

[2] Behrens, A.M., Choobineh, F., *Can Economic Uncertainty Always be Described by Randomness?*, International Industrial Engineering Conference & Societies' Manufacturing and Productivity Symposium, Institute of Industrial Engineers, 1989

[3] Box, G.E.P., Draper, N.R., Empirical Model-Building and Response Surfaces, John Wiley & Sons, Inc., New York, NY, 1987.

[4] Box, G.E.P., Hunter, W.G., Hunter, J.S., Statistics for Experimenters, John Wiley & Sons, Inc., New York, NY, 1978.

[5] Chen, W., et al., *Integration of Response Surface Method with the Compromise Decision Support Problem in Developing a General Robust Design Procedure*, Advances in Design Automation (Azarm, S., et al. Eds.), Vol. 82-2, ASME, New York, 1995

[6] Cheng, B., Titterington, D.M., *Neural Networks: A Review from a Statistical Perspective*, Statistical Science, Vol. 9, No. 1, 2-54, 1994

[7] DeLaurentis, D.A., Mavris, D.N., and Schrage, D.P., *An IPPD Approach to the Preliminary Design Optimization of an HSCT using Design of Experiments*, 20th ICAS Congress, Sorrento, Italy, September, 1996.

[8] DeLaurentis, D.A., Calise, A., Schrage, D.P., and Mavris, D.N., *Integrating Guidance Optimization into Vehicle Design Via Singular Perturbation Theory and Statistical Methods*, 6th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Bellevue, WA, September, 1996.

[9] Dieter, G.E., Engineering Design, A Materials and Processing Approach, 2nd Edition, McGraw-Hill, Inc., New York, NY, 1991.

[10] Fabrycky, W.J., Blanchard, B.S., Life-Cycle Cost and Economic Analysis, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1991.

[11] Fox, E.P., *The Pratt & Whitney Probabilistic Design System*, 35th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Hilton Head, SC, April, 1994

[12] Galloway, T.L., and Mavris, D.N., Aircraft Life Cycle Cost Analysis (ALCCA) Program,

NASA Ames Research Center, September 1993.

[13] Giunta, A.A., et al, *Wing Design for a High-Speed Civil Transport Using a Design of Experiments Methodology*, 6th AIAA/ USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Bellevue, WA, September, 1996

[14] Haugen, E.B., Probabilistic Mechanical Design, Wiley-Interscience, New York, NY, 1980.

[15] Huber, K.-P., Berthold, M. R., Szczerbicka, H., *Analysis of Simulation Models with Fuzzy Graph based Metamodeling*, Performance Evaluation, Vol. 27&28, North-Holland, October 1996

[16] Kaufman, M., et al, *Variable-Complexity Response Surface Approximations for Wing Structural Weight in HSCT Design*, 34th Aerospace Sciences Meeting & Exhibit, Reno, NV, January, 1996.

[17] Kleijnen, J.P.C., Statistical Techniques in Simulation, Part I, Marcel Dekker, Inc., New York, NY, 1974.

[18] Kleijnen, J.P.C., Statistical Tools for Simulation Practitioners, Marcel Dekker, Inc., New York, NY, 1987.

[19] Kusiak, A., Concurrent Engineering, John Wiley & Sons, Inc., New York, NY, 1993.

[20] Mavris, D.N., Bandte, O., Schrage, D.P., *Economic Uncertainty Assessment of an HSCT Using a Combined Design of Experiments/ Monte Carlo Simulation Approach*, 17th Annual Conference of the International Society of Parametric Analysts, San Diego, CA, June, 1995.

[21] Mavris, D.N., Bandte, O., *Effect of Mission Requirements on the Economic Robustness of an HSCT*, 18th Annual Conference of the International Society of Parametric Analysts, Cannes, France, June, 1996.

[22] Mavris, D.N., Bandte, O., Schrage, D., *Application of Probabilistic Methods for the Determination of an Economically Robust HSCT Configuration*, AIAA/USAF/NASA/ ISSMO Multidisciplinary Analysis and Optimization Conference, Bellevue, WA, September, 1996.

[23] McCullers, L.A., Flight Optimization System, Computer Program and Users Guide, Version 5.7, NASA Langley Research Center, Hampton, VA, December 1994.

[24] Neter, J., Wassermann, W., Kutner, M.H., Applied Linear Statistical Models, Richard D. Irwing, Inc., Homewell, IL, 1990.

[25] Sacks, J., Schiller, S.B., Welch, W.J., *Design for Computer Experiments*, Technometrics, Vol. 31, No. 1, February, 1989

[26] Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P., *Design and Analysis of Computer Experiments*, Statistical Science, Vol. 4, No. 4, February, 1989

[27] Siddall, J.N., Probabilistic Engineering Design, Marcel Dekker, Inc., New York, NY, 1983.

[28] Southwest Research Institute, FPI User's and Theoretical Manual, San Antonio, TX, 1995.

[29] Unal, R., Stanley, D.O., Joyner, C.R., *Parameter Model Building and Design Optimization Using Response Surface Methods*, Journal of Parametrics, Washington, DC, May, 1994

[30] Welch, W.J., Buck, R.J., Sacks, J., Wynn, H.P., Mitchell, T.J., Morris, M.D., *Screening, Predicting, and Computer Experiments*, Technometrics, Vol. 34, No. 1, February, 1992

[31] Wu, Y.-T., Burnside, O.H., Dominguez, J., *Efficient Probabilistic Fracture Mechanics Analysis*, 4th International Conference on Numerical Methods in Fracture Mechanics, Pineridge Press, Swansea, U.K., May, 1987